Model and priors

Our model specifies a logistic curve that fits our outcome measure. Out outcome measure is the porportion of participants that are reported to understand each word. This logistic curve is characterised by three parameters: *asymptote* ($asym$, the value at which the outcome *pleateaus*), *steepness*, ($steep$, how fast the curve grows), and *mid-point* ($mid - point$, the age at which the curve reaches its maximum steepness). We identify the *mid-point* of a word as its age of acquisition, and thus us the parameter of interest. Out hypothesis is that the language a word belongs to (non-dominant vs. dominant), together with its cognate status (non-cognate vs. cognate) and frequency, affects the position of its mid-point. We assume that the mid-point is generated from a zero-one beta regression model that includes the aforementioned predictors. A beta regression model helps us model data that include many values at boundary, as ours are. We also included random intercepts for each *translation equivalent* (TE). We set informed priors for *asymptotes*, *steepnesses*, and *mid-points* according to Wordbank data, and weak priors the the coefficients of the beta-regression model.

$$
\begin{aligned}
proportion_{ij} &\sim \frac{asym}{1 + e^{(mid_i - age) \times steep}} \\
asym &\sim Normal(0.79, 0.1) \\
steep &\sim Normal(1.76, 0.8) \\
mid &\sim Beta(\mu_i, \phi) \\
\mu_i &= \beta_0 + TE_{0i} + \beta_1 \times dominance_i + \beta_2 \times cognateness_i \\
\beta_0 &\sim Normal(4.50, 1) \\
\beta_1 &\sim Normal(0, 10) \\
\beta_2 &\sim Normal(0, 10) \\
\phi &\sim Normal(1.5, 1)
\end{aligned}
$$